Alain Polguère*
(Nancy)

LEXICOGRAPHIE DES DICTIONNAIRES VIRTUELS

Résumé. La lexicographie est traditionnellement conçue comme discipline de construction de dictionnaires. Or, si le dictionnaire représente bien une forme exploitable de visualisation du lexique, il n'en est aucunement une image structurale. Le lexique est en effet, formellement, un graphe immensément riche de connexions entre entités lexicales alors que le dictionnaire, dans sa réalité formelle, n'est qu'un « texte ». La prochaine étape d'évolution de la lexicographie doit être celle d'une lexicographie des dictionnaires virtuels, qui vise en tout premier lieu la construction de structures informationnelles à l'image de la structure véritable (ou plausible) des lexiques, structures à partir desquelles peuvent être générées des visualisations aussi bien diagrammatiques que textuelles de l'information lexicale. Nous présentons cette lexicographie à partir de l'entreprise de de construction du *Réseau Lexical du Français* (RLF), qui est ancrée dans la Lexicologie Explicative et Combinatoire (LEC) bien qu'elle ne vise pas la rédaction proprement dite des dictionnaires postulés par la LEC.

Abstract. One generally sees dictionary writing as being the essence of lexicography. Though dictionaries are indeed useful means of visualizing lexicons, they are by no means their structural representation. This is due to the fact that a lexicon is, formally, a immensely rich graph of connected lexical entities while a dictionary is, in essence, nothing but a "text". The next step in the evolution of lexicography will be toward the lexicography of virtual dictionaries, whose first aim is the building of informational structures analogous to the actual (or plausible) structure of lexicons; from such structures, both diagrammatic and textual visualizations of lexical information can be generated. We introduce this new lexicography, exemplifying it with the construction of the *Réseau Lexical du Français* (RLF, 'French Lexical Network'). We show that, while deeply rooted in Explanatory Combinatorial Lexicology (ECL), this enterprise does not target the actual writing of the type of dictionaries postulated by ECL, as a theoretical and descriptive framework.

Mots-clés: Dictionnaire Explicatif et Combinatoire (DEC), dictionnaire virtuel, Réseau Lexical du Français (RLF), système lexical, fonction lexicale

Keywords: Explanatory-Combinatorial Dictionary (ECD), virtual dictionary, 'French Lexical Network' (RLF), lexical system, lexical function

1. De la rédaction de dictionnaires à la construction de lexiques

La lexicographie est traditionnellement conçue comme discipline de construction de dictionnaires. Mais qu'est-ce qu'un dictionnaire? Il a maintes fois été

* Université de Lorraine, CNRS, ATILF, UMR 7118. 44 avenue de la Libération, BP 30687 Nancy cedex, F-54063, France ; alain.polguere@univ-lorraine.fr

souligné qu'il existe une très grande variété d'objets appelés *dictionnaires* : dictionnaires de langue, dictionnaires monolingues, dictionnaires multilingues, dictionnaires d'apprentissage, dictionnaires historiques, etc., jusqu'aux ouvrages appelés *dictionnaires* uniquement parce qu'ils se structurent comme des collections d'articles relativement courts portant sur des notions données (dictionnaire philosophique, dictionnaire des rêves, etc.). Si nous nous en tenons aux dictionnaires « véritables », ceux qui visent la description des unités du lexique d'une ou plusieurs langues, on constate que ces objets descriptifs possèdent tous les propriétés structurales suivantes, sous leur forme imprimée aussi bien que sous leur forme électronique :

1. ce sont des textes constitués d'une succession de très courts sous-textes (classés alphabétiquement, ou par tout autre moyen permettant un ordonnancement et une récupération de l'information);
2. ils ont une structure rigide et la structure de stockage de l'information par les lexicographes est en très grande partie identique à celle de l'information rendue disponible aux utilisateurs.

De ces deux caractéristiques formelles fondamentales découlent deux limitations inhérentes aux dictionnaires: une limitation théorique et une limitation pratique.

Sur le plan théorique, les dictionnaires, dans leur réalité textuelle, représentent bien une forme de visualisation du lexique. Ils n'en sont cependant aucunement une image structurale. Le lexique est en effet, formellement, un graphe immensément riche de connexions entre entités lexicales alors que les dictionnaires, dans leur réalité formelle, ne sont que des « textes ».

Sur le plan pratique, du fait de leur structure rigide prédéterminée, les dictionnaires visent tous un type d'utilisateur donné, plus ou moins précisément identifié. Cela pousse les maisons d'édition de dictionnaires à changer sans cesse la gamme de leurs produits lexicographiques, déclinant de façon multiple la description du lexique avec des nomenclatures variables, des articles aux contenus variables et adaptés, etc.

Ne serait-ce que pour ces deux raisons, la prochaine étape d'évolution de la lexicographie, en tant que discipline, doit être celle de la lexicographie des dictionnaires virtuels: une lexicographie qui vise en premier lieu la construction d'une structure informationnelle à l'image de la structure véritable (ou plausible) du lexique et à partir de laquelle peuvent être générées des visualisations de l'information lexicale aussi bien diagrammatiques — c'est-à-dire, sous forme de diagrammes, au sens de C. S. Peirce [1] — que textuelles. Nous présentons cette lexicographie à partir de l'entreprise de de construction du *Réseau Lexical du Français* ou RLF, qui est ancrée dans la Lexicologie Explicative et Combinatoire ou LEC (Mel'čuk et al. 1995) bien qu'elle ne consiste pas en la rédaction des dictionnaires postulés par la LEC.

[1] Voir, par exemple, (Farias, Queiroz 2006).

Attention: ce que nous venons de dire sur les dictionnaires s'applique aussi bien aux dictionnaires commerciaux qu'aux dictionnaires dits *théoriques*. Ces derniers sont aussi des textes et ils visent un type d'utilisateur unique : le chercheur en linguistique. Un utilisateur important est ainsi laissé de côté : la machine. Même si le dictionnaire théorique s'annonce comme formel, il n'est dans la pratique que rédigé en « pseudo-code » et est du texte, rien d'autre que du texte: de l'information non calculable en dépit des formalismes et formules qui peuvent y être employés. Pour prendre le cas du *Dictionnaire Explicatif et Combinatoire* (DEC), puisqu'il s'agit bien là de la lexicographie qui est notre centre d'attention dans cet article, la notion même d'informatisation d'un DEC est une notion sans lendemain, car un DEC est un texte : on ne peut que le stocker sous forme numérique, structuré par des balises. Cela permet déjà de faire énormément de choses, qui seraient impossibles avec un dictionnaire ordinaire, comme l'a montré l'expérience de l'informatisation du *Trésor de la Langue Française* (Dendien, Pierrel 2003). Mais il s'agit d'une expérience qui atteint vite ses limites: le plus beau dictionnaire du monde ne donne que ce qu'il a.

Pour faire notre démonstration de l'importance d'évoluer vers une lexicographie des dictionnaires virtuels, nous procédons en deux étapes. Tout d'abord, section 2, nous faisons un bref retour « historique » sur l'application de la LEC à la description du français. Ensuite, section 3, nous montrons comment la LEC peut être appliquée à la mise en œuvre d'une lexicographie des dictionnaires virtuels.

Nous offrons à Igor Mel'čuk, pour ce recueil d'hommages, un type de texte qu'il n'aime pas lire. Mais les instructions données aux rédacteurs étant de faire court, nous nous consolons en nous disant qu'il ne souffrira que brièvement de cette littérature. Dans notre esprit, le véritable hommage que nous rendons à Igor, avec émotion, c'est le travail lexicographique dont il est question ici et dont nous espérons pouvoir éclairer la finalité.

2. Retour sur la pratique de la Lexicographie Explicative et Combinatoire du français

En avril 2012 a débuté au laboratoire ATILF de Nancy un projet lexicographique de construction d'une ressource lexicale appelée *Réseau Lexical du Français* ou RLF (Lux-Pogodalla, Polguère 2011). Le RLF est entièrement conçu selon les principes de la Lexicologie et Lexicographie Explicatives et Combinatoires — LEC et LgEC [2]. Ces principes sont détaillés dans de nombreuses publications — dont (Mel'čuk 2006) — et il est inutile de les résumer ici. Il est par contre utile de situer l'entreprise de construction du RLF dans la perspective de

[2] Nous nous permettons d'introduire le sigle LgEC, car nous avons fréquemment à faire référence spécifiquement à l'activité lexicographique (effectuée selon les principes de la LEC).

l'évolution de la LgEC du français telle qu'elle a été pratiquée au cours des trente dernières années. Nous nous focalisons sur la description du français, car il s'agit de celle que nous connaissons le mieux et sur laquelle nous avons un point de vue à la fois d'observateur et de contributeur.

Originellement, à partir des années 80, la LgEC du français a visé avant tout la production de volumes du *Dictionaire Explicatif et Combinatoire du Français Contemporain* ou DECFC. Comme on le sait, quatre volumes on été publiés à Montréal sous la direction d'I. Mel'čuk, de 1984 à 1999 (Mel'čuk et al. 1984—1999)[3]. La rédaction d'articles du DECFC visait des buts similaires à la rédaction d'articles scientifiques, le DECFC étant considéré comme un dictionnaire théorique — cf. le sous-titre de tous les volumes publiés: ***Recherches*** *lexico-sémantiques*. La contribution du travail sur le DECFC à l'étude du lexique français et au développement des notions fondamentales de la LEC a été très importante. Au-delà des apports théoriques, il convient de mentionner que le travail effectué sur le DECFC a été le principal support de formation en lexicologie théorique et descriptive de plusieurs générations d'étudiants et collaborateurs (dont votre serviteur).

Si le projet lexicographique du DECFC s'est avéré novateur de par ses fondements et sa finalité théorique, il a peu innové sur le plan méthodologique. On peut même dire qu'il était très en deçà de l'état de l'art en lexicographie « traditionnelle et commerciale », notamment pour ce qui est de l'outillage logiciel et du recours aux corpus linguistiques. Ces derniers ont bien entendu été mis à l'œuvre, mais sans qu'une méthode précise de consultation et d'exploitation des données ne soit élaborée. La taille des corpus utilisés était médiocre et les méthodes d'accès aux données ne présentaient aucune caractéristique dont on puisse s'enorgueillir. Quant à l'outillage logiciel pour l'édition lexicographique proprement dite, il était presque inexistant, puisque les articles étaient banalement rédigés au moyen d'un traitement de texte commercial standard (que nous ne nommerons pas), puis proprement reformatés en phase de publication à l'aide d'un logiciel d'édition professionnel. Les praticiens de la LgEC ont toujours été conscients de l'écart entre la complexité du modèle lexical que représente un DEC et l'incapacité d'un traitement de texte ordinaire d'apporter quelque support que ce soit pour gérer cette complexité. Plusieurs tentatives ont de ce fait eu lieu pour élaborer un outil dédié à l'édition lexicographique des DEC, sans que cela débouche sur la construction d'un logiciel véritablement exploitable (Décary, Lapalme 1990; Sérasset 1998). Nous avons toujours été très sceptique quant à la viabilité d'une informatisation de la rédaction des DEC et cet échec de l'entreprise d'informatisation des DEC a été plus une confirmation de notre intuition initiale

[3] Le premier volume du DECFC a été publié la même année que le DEC russe (Mel'čuk, Zholkovsky 1984).

qu'une déception véritable. Pour nous, ce sur quoi il fallait travailler en premier lieu, d'un point de vue méthodologique, a toujours été la refonte de la structure de données et non l'outillage informatique pour l'édition d'un texte lexicographique.

Le projet de construction de la base lexicale *DiCo* (pour *dictionnaire de combinatoire*) a débuté dans les années 90 sous la forme d'une collaboration entre le présent auteur et Igor Mel'čuk (Jousse, Polguère 2005; Mel'čuk, Polguère 2006). Il s'agissait de poursuivre le travail effectué sur le français dans le cadre de la LgEC, mais avec une structure de données plus appropriée au traitement automatique, qui puisse déboucher à la fois sur l'implantation d'une véritable lexicographie informatisée et sur des exploitations directes de la ressource en traitement automatique de la langue. Trois caractéristiques du DiCo permettaient cela.

Premièrement, le DiCo simplifie le problème lexicographique en remplaçant la définition lexicographique par une spécification de la structure actancielle de chaque lexie-vedette, accompagnée d'un étiquetage sémantique de celle-ci. D'un point de vue sémantique, le DiCo est donc un « sous-DEC », même si le travail de construction d'une hiérarchie d'étiquettes sémantiques est original en regard de la pratique classique de la LgEC et présente un intérêt certain pour la structuration et le traitement de l'information lexicale (Polguère 2011).

Deuxièmement, le DiCo a dès le départ été construit à l'aide de logiciels permettant la gestion de l'information structurée : tout d'abord, avec un logiciel maison (Polguère 1994), remplacé ensuite par un logiciel professionnel de gestion de base de données lorsque la base lexicographique a atteint une dimension requérant le recours à un environnement informatique véritablement robuste.

Troisièmement, le contenu informationnel du DiCo a été dès le départ structuré en anticipant la nécessité de son traitement informatique. Chaque fiche DiCo décrivant une lexie-vedette a été explicitement subdivisée en champs dédiés à un type bien défini de caractérisation de la lexie-vedette et le texte de chaque champ a été structuré par un balisage et des conventions d'écriture permettant une analyse et une restructuration automatique de l'information stockée.

Même si le DiCo peut être vu comme une simplification du DEC, il présente par rapport à ce dernier un double enrichissement. Tout d'abord, la structuration formelle de l'information lexicographique revient de fait à un enrichissement de cette dernière. Ensuite, l'encodage des relations paradigmatiques et syntagmatiques au moyen des fonctions lexicales (Mel'čuk 1996), qui est au cœur de la structuration de l'information lexicale dans un modèle produit par la LEC, a été enrichi par un encodage grand public, visant une autre gamme d'utilisateurs potentiels des modèles lexicaux explicatifs et combinatoires: les apprenants et les enseignants du français, non spécialistes de la LEC. Il s'agit des formules dites *formules de vulgarisation*, qui paraphrasent en un français élémentaire le contenu sémantico-syntaxique de chaque lien de fonction lexicale standard ou semi-standard.

Le travail sur le DiCo a permis la modélisation d'un nombre significatif de lexies du français [4]. Cependant, la base construite, tout comme le DECFC, est en réalité un échantillon de dictionnaire, si on la compare à un dictionnaire de langue standard. Cela s'explique par deux faits, dont le second nous intéresse ici tout particulièrement. Premièrement, il n'a jamais été possible de trouver un financement direct pour faire passer le travail de construction du DiCo à la vitesse supérieure avec la mise en place d'une véritable équipe de rédaction lexicographique. Les fiches lexicographiques ont été construites par les deux chercheurs impliqués, avec la collaboration d'étudiants pour lesquels l'activité lexicographique de construction de fiches DiCo servait les mêmes finalités de formation que la rédaction des articles du DECFC. Deuxièmement, même si l'on a recours à une approche plus appareillée et formalisée comme celle du DiCo, les modèles lexicaux postulés par la LEC sont d'une telle complexité qu'ils ne peuvent tout simplement pas être développés sur une structure de dictionnaire au-delà d'une certaine masse d'information construite, que nous évaluons par expérience à environ 1 000 lexies. Passé cette limite, l'introduction dans la base de données d'une nouvelle lexie génère plus de travail d'insertion (connexion aux données déjà présentes) que la recherche des propriétés linguistiques de la lexie en question.

Cela s'explique par le fait que la véritable structure d'un modèle explicatif et combinatoire du lexique de la langue est un réseau lexical d'un type particulier — un système lexical (Polguère 2009) — et non un texte. Chercher à construire et éditer un système lexical sous la forme d'un texte est irréaliste et, plutôt que de développer des outils informatiques pour effectuer cela, la solution la plus logique est de construire directement un système lexical sous la forme d'un graphe. Les textes lexicographiques ne sont alors que des produits dérivés potentiels de ce graphe: ce sont des **dictionnaires virtuels**.

Une « démonstration de concept » pour la production de données dictionnairiques à partir d'une structure de données relationnelles où l'information lexicographique est atomisée a été effectuée à l'Observatoire de linguistique Sens-Texte (OLST) de l'Université de Montréal. Dans un premier temps, une collaboration avec le laboratoire Lattice du CNRS a permis d'effectuer la mise en ligne d'une version SQL (*Standard Query Language*) du DiCo, où toute l'information lexicographique textuelle contenue dans les fiches DiCo est automatiquement « saucissonnée » et répartie dans un ensemble de tables SQL formant une base de données relationnelle (Steinlin et al. 2005). Cette compilation est rendue relativement aisée par le fait que le DiCo a justement été conçu, dans sa formalisation, pour autoriser ce type de traitement. Une interface de consultation, le

[4] Pour les fiches lexicographiques de statut « 0 » (finalisées) et « 1 » (préfinalisées): 1129 lexies, regroupées en 539 vocables, connectées par 26 112 liens de fonctions lexicales.

Dicouèbe [5], permet un accès par formulaire (ou par exécution directe de requêtes SQL) aux données tabulaires générées à partir des données textuelles du DiCo. Une seconde étape — la réalisation de l'environnement DiCopop — a permis d'expérimenter la génération automatique d'articles dictionnaires grand public à partir des données du DiCouèbe, articles structurés sur le modèle des articles du *Lexique Actif du Français* (Mel'čuk, Polguère 2007), avec les apprenants et les enseignants de langue comme utilisateurs de référence. Il n'existe malheureusement pas de publication présentant l'expérience du DiCopop, mais ce dernier est accessible en ligne sur le site de l'OLST et est toujours une source de consultation très pratique des données du DiCo [6]. Les articles lexicographiques du DiCopop sont générés à la volée, à chaque consultation par un visiteur du site. De ce point de vue, l'expérience de construction du DiCopop a été la première étape concrète de génération automatique d'articles lexicographiques dans le cadre de la LgEC du français et, à notre connaissance, de la LgEC dans son ensemble.

Comme nous l'avons mentionné en début de section, l'évolution la plus récente de la LgEC du français est le démarrage de la construction du *Réseau Lexical du Français* ou RLF. C'est le sujet que nous allons maintenant aborder.

3. Lexicographie Explicative et Combinatoire à la sauce RLF

La construction du RLF a débuté dans le cadre d'un projet de trois ans appelé *RELIEF*. Faute de place, nous ne ferons pas ici de présentation de RELIEF; pour une introduction générale, on pourra se reporter à (Lux-Pogodalla, Polguère 2011) ainsi qu'à la page web du projet, sur le site du laboratoire ATILF [7]. Nous nous concentrons, dans cette section, sur deux points cruciaux pour la compréhension de ce qu'est la lexicographie des dictionnaires virtuels, telle que mise en pratique dans le cadre de la construction du RLF : la structure formelle du RLF (3.1) et le processus d'édition lexicographique (3.2). Mentionnons que RELIEF achève à peine sa première année d'exécution au moment où nous écrivons ces lignes. Par conséquent, la perspective donnée ici est partielle et limitée à l'état d'avancement du travail.

3.1. Le RLF: un lexique à structure de graphe et non un dictionnaire

Dès le départ, le projet de construction du RLF a été conçu de telle sorte qu'il respecte les trois contraintes fondamentales suivantes :

---

[5] L'interface DiCouèbe est accessible à l'adresse http://olst.ling.umontreal.ca/dicouebe. Avant de l'utiliser, il est fortement recommandé de télécharger et lire le document descriptif (Jousse, Polguère 2005).

[6] http://olst.ling.umontreal.ca/dicopop.

[7] http://www.atilf.fr/spip.php?article908.

1. Le RLF doit être construit comme un système lexical — un graphe lexical dont les arcs sont tissés (en premier lieu) par des liens de fonctions lexicales (Polguère 2009).
2. Le RLF doit avoir un contenu informationnel équivalent, au moins, à celui d'un DEC.
3. La méthodologie de construction du RLF repose sur le recours à un éditeur lexicographique dédié qui, bien que ne gérant pas la construction d'un dictionnaire, supporte une méthode de description des lexiques qui relève entièrement de la LgEC.

Insistons sur ce dernier point. La réalisation du RLF est un travail lexicographique véritable: la lexicographie des dictionnaires virtuelles n'est pas une lexicographie virtuelle [8].

Les trois contraintes ci-dessus ont été clairement posées dès la conception du projet et c'est à partir de celles-ci que s'est engagée une collaboration entre les membres de l'équipe lexicographique du RLF et Nabil Gader (MVS, Saint-Dié), le développeur de l'éditeur lexicographique. Cet éditeur, appelé *Dicet*, emprunte une partie de l'architecture du logiciel de saisie rédactionnelle Dixit de l'entreprise MVS [9], même s'il convient de considérer que Dicet est une réalisation tout à fait originale sur le plan informatique.

Dicet permet de construire le modèle du lexique directement en alimentant un ensemble très riche de tables SQL, où chaque entité linguistique ajoutée au lexique — unité lexicale, vocable, fonction lexicale, application de fonction lexicale, etc. — possède une identification propre, qui la réifie en tant que nœud du graphe lexical. Pour aider le lecteur à comprendre en quoi la construction des données du RLF est fondamentalement différente de la construction d'articles de dictionnaire, le plus simple est sans doute de prendre un exemple concret.

3.2. Illustration du processus d'édition lexicographique

La lexie ABOYER **I** [*Son chien aboie.*] existe de façon unique dans le RLF par un numéro d'identification stocké dans une des tables SQL de la base — par exemple, le numéro 30524. À chaque fois que le lexicographe associe une propriété à cette lexie ou qu'il la « mentionne » dans le cadre de la description d'une autre lexie, il pointe dans la réalité vers ce code unique, et cela vaut pour toute lexie ou, plus généralement, toute entité lexicale du RLF.

---

[8] Les entreprises qui visent à construire automatiquement des ressources lexicales à partir de ressources autres (Sagot, Fišer 2008; Alonso et al. 2011) relèvent, elles, d'une lexicographie virtuelle.

[9] http://www.mvs.fr/pdf/MVS_Dixit.pdf.

Dans une approche traditionnelle de construction de dictionnaire, le lexicographe indiquerait que les lexies FURIEUSEMENT et FÉROCEMENT sont deux **Magn** (collocatifs intensificateurs) de ABOYER **I** en ajoutant dans l'article de cette lexie un bloc de texte du type suivant, où Intensément est la formule de vulgarisation de **Magn**(ABOYER **I**):

Intensément
Magn: *furieusement*; *férocement*

Le lexicographe du RLF ne va rien « écrire » pour effectuer la même tâche lexicographique. Il va plutôt tisser avec Dicet le lien de fonction lexicale mis en jeu ici (entre la lexie-vedette et l'ensemble de ses deux intensificateurs), tel qu'on le voit dans la Figure 1 ci-dessous.

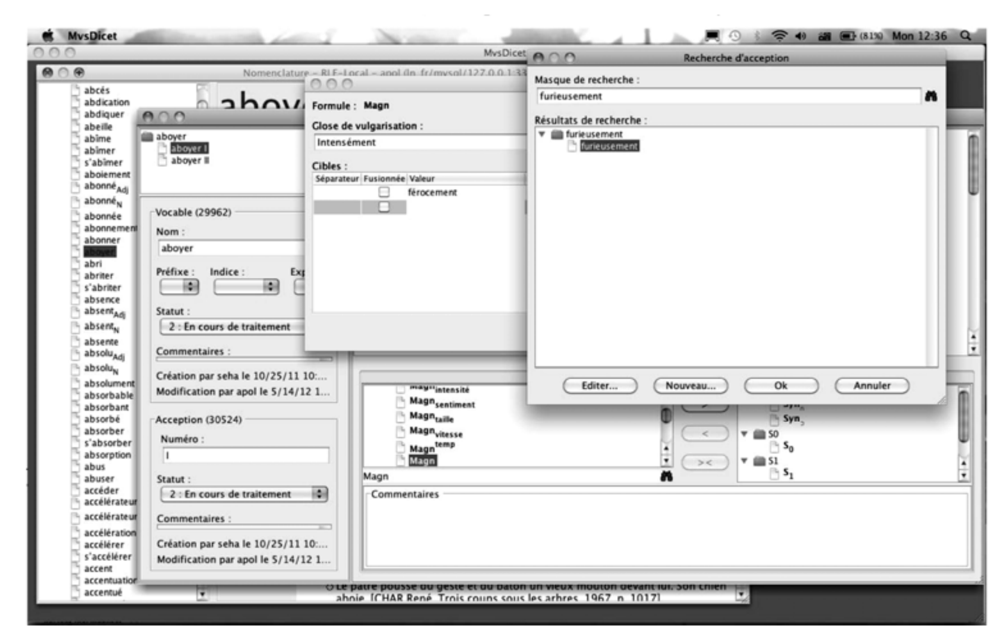


Figure 1: Sélection d'une valeur de **Magn** pour une lexie-vedette dans Dicet

Dans cette figure, apparaissent en superposition quatre fenêtres qui correspondent à quatre étapes de tissage du lien ; il s'agit (de l'arrière-plan au premier plan de la figure) :

1. de la fenêtre de visualisation de la nomenclature du RLF, qui permet notamment l'accès à la description d'un vocable et des lexies qu'il regroupe ;
2. de la fenêtre de visualisation de la fiche lexicographique d'une lexie-vedette, qui permet notamment l'édition du champ de fonctions lexicales (ici, sélection de **Magn** pour le tissage des liens correspondants) ;

3. de la fenêtre de tissage d'un lien de fonction lexicale, avec sélection de la formule de vulgarisation et entrée des cibles lexicales ;
4. de la fenêtre de recherche d'une cible de fonction lexicale dans le RLF.

Il faut noter que cette dernière fenêtre donne accès non seulement à la sélection d'une lexie du RLF, mais aussi à sa consultation et à son édition lexicographique. De plus, elle permet de créer à la volée une nouvelle lexie (nouvelle acception d'un vocable existant ou acception d'un nouveau vocable) si la cible lexicale désirée n'existe pas déjà dans le RLF. Le lexicographe fait, dans un tel cas, une description minimale provisoire de la lexie créée à la volée : partie du discours, genre (pour les noms) et exemple(s) lexicographique(s). On voit qu'il ne s'agit donc pas simplement ici de tisser des liens de fonctions lexicales, mais aussi, si besoin est, de lancer de façon récursive une opération de description lexicographique d'une ou plusieurs lexies à l'intérieur du processus de description d'une lexie-vedette. Cette façon de procéder, par création d'entrées et fiches lexicographiques à la volée, est particulièrement adaptée à un travail lexicographique grandeur nature, à l'échelle de l'ensemble du lexique de la langue et en mettant en jeu une véritable équipe lexicographique.

Un lien de fonction lexicale tissé au moyen de Dicet peut être ultérieurement supprimé, renommé (si l'on juge que l'on est en présence d'une autre fonction lexicale) ou précisé (par une autre formule de vulgarisation, des contraintes additionnelles, etc.). Le travail du lexicographe consiste à se concentrer sur l'identification et la caractérisation d'une propriété de combinatoire spécifique de la lexie-vedette et rien d'autre.

À partir de l'information que le lexicographe introduit dans le RLF (par création d'entités informationnelles et tissage de liens), l'éditeur Dicet donne un retour textuel dans une « vue-article », où la zone de fonctions lexicales d'un article de DEC virtuel est automatiquement générée simultanément à l'introduction de données dans la base — cf. Figure 2.

Fonctions lexicales

$\mathbf{Syn}_{\cap}$ glapir; hurler I

$\mathbf{Syn}_{\supset}$ japper

$\mathbf{S_0}$ aboiement I

Nom pour $1
$\mathbf{S_1}$ chien I

Intensément
Magn férocement; furieusement

Peu intensément
AntiMagn faiblement, //japper; sourdement

Figure 2: Vue-article du champ de fonctions lexicales de ABOYER **I** dans le RLF [10]

[10] Le symbole « $1 » utilisée dans la formule de vulgarisation de **$\mathbf{S_1}$** vaut pour la variable identifiant le premier actant de la lexie-vedette — généralement, *X*. Cette variable

La vue-article présentée ci-dessus n'est qu'une visualisation **textuelle** du contenu du RLF. La visualisation la plus proche de la réalité informationnelle est de nature **diagrammatique**, analogue à la structure du lexique: c'est un graphe où apparaissent les liens existant entre les différentes entités du RLF. Le graphe de la Figure 3 ci-dessous a été généré automatiquement à partir des données du RLF ; il présente les liens de fonctions lexicales dont la lexie ABOYER **I** est la source ou la cible au moment où nous écrivons.

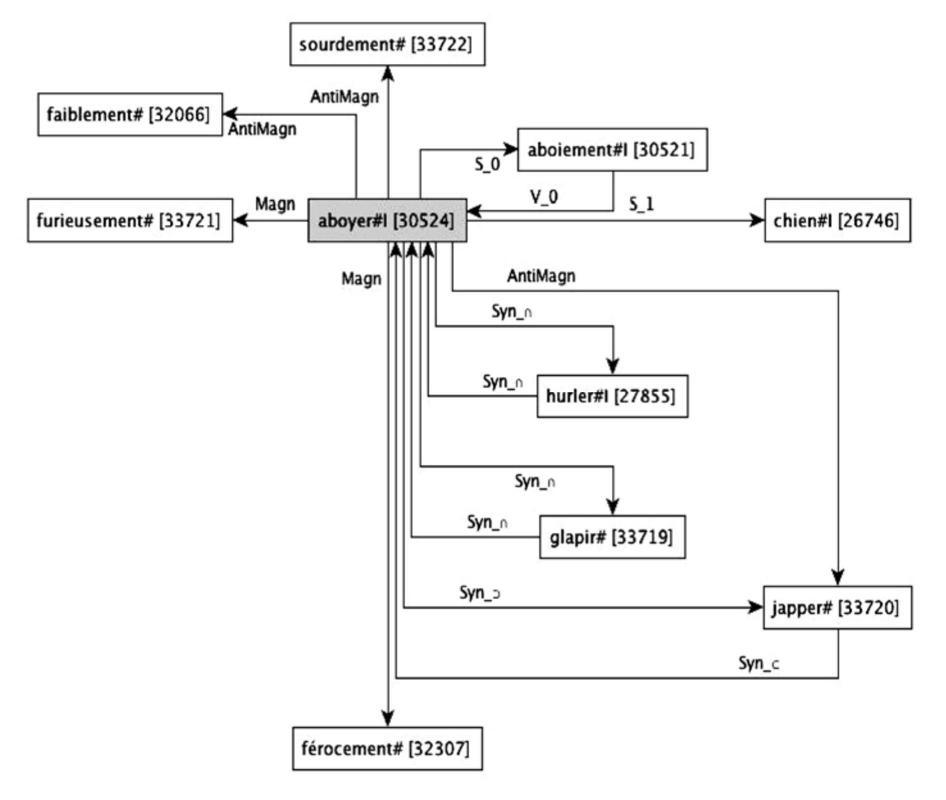


Figure 3: Visualisation graphique des liens dont ABOYER **I** est la source ou la cible

Même si la structure de données correspondant à la zone de fonctions lexicales de la lexie-vedette ABOYER **I** est beaucoup plus riche que ce que révèle la Figure 3, cette dernière reflète, bien mieux que la Figure 2, la véritable nature de graphe du RLF. C'est parce que le RLF est un graphe, et non un texte, que de multiples

apparaitra dans les articles à la place de « $1 » une fois que le champ de définition (qui inclut la spécification de la forme propositionnelle de la lexie-vedette) aura été implémenté dans Dicet.

opérations lexicographiques (vérification, hypothèses, raisonnements analogiques, etc.) sont rendues possibles, et cela, **durant le processus de construction du RLF**.

Il est très important de comprendre quelles sont les implications de cette façon de construire la description lexicale. Il ne s'agit pas simplement d'épargner au lexicographe le travail fastidieux de formatage du texte d'un article, ce que l'écriture d'un dictionnaire sous forme de base XML avec application de feuilles de style permet facilement de faire. Ce qui est remarquable ici, c'est qu'une modification ponctuelle dans la description d'une lexie n'entraîne aucune tâche de mise à jour dans l'ensemble de la base. Que ce soit pour faire référence à des lexies, à des fonctions lexicales, à des formules de vulgarisation, etc., le lexicographe ne fait que pointer via Dicet sur des identificateurs d'entités lexicographiques. La façon dont ces entités apparaissent ensuite dans une vue-article, dans des menus de Dicet, etc., est entièrement conditionnée par la modélisation qu'en fait le lexicographe. Dicet permet non seulement d'éditer les articles de lexies, mais il permet aussi dans sa version actuelle :

1. l'édition de l'ensemble des caractéristiques grammaticales utilisées dans les articles ;
2. l'édition des descriptions des fonctions lexicales, en tant que métalexies.

Dicet possède ainsi la « connaissance » (du moins une connaissance encore partielle) du système des fonctions lexicales. Chaque fonction lexicale peut être éditée et sa description modifiée (nom exact, formules de vulgarisation possibles, etc.), les modifications devenant immédiatement effectives dans l'ensemble de la base et ayant un effet direct sur les différentes visualisations permises par Dicet.

À ce jour, seuls les champs de caractéristiques grammaticales et de fonctions lexicales ont été entièrement implantés; les données des autres champs sont pour l'instant saisies et stockées de façon conventionnelle, comme il était fait dans le DiCo. L'ordre chronologique planifié pour l'implantation des autres champs est:

1. exemples (en cours d'élaboration) ;
2. définition (avec étiquetage sémantique, forme propositionnelle et paraphrase définitionnelle proprement dite) ;
3. phraséologie (connections aux locutions formées à partir de la lexie-vedette) ;
4. régime syntaxique ;
5. nota bene.

Les informations de tous les champs sont bien entendu conçues pour être connectables. Ainsi, l'information qui sera formalisée par Dicet dans le champ de définition d'une lexie-vedette sera automatiquement connectée à l'édition, notamment, du champ de fonctions lexicales : une structure actancielle donnée conditionne la

possibilité ou l'impossibilité de certains liens de fonctions lexicales, une étiquette sémantique donnée conditionne le choix de certaines formules de vulgarisation pour une application d'une fonction lexicale donnée, etc.

Les répercussions méthodologiques et pratiques de l'édition d'un RLF, et non d'un DEC, sont infinies, car il n'y a tout simplement pas de limite à l'exploitation que l'on peut faire de la structuration en graphe des données. Faute de place, nous n'avons malheureusement pu donner ici qu'une illustration très ponctuelle de cela. Nous nous permettons néanmoins d'avancer l'hypothèse que le passage d'une représentation textuelle de la structure du lexique à une représentation en graphe entraîne un gain (théorique et pratique) pour la LEC/LgLEC analogue au passage d'une représentation linéaire du sens — fondée sur le langage de la logique de prédicats — à une représentation sous forme de réseau sémantique pour l'étude de la sémantique.

4. Récapitulatif

Le point essentiel que nous avons présenté ici est le suivant: la lexicographie — et, plus spécifiquement, la LgEC — ne doit pas avoir pour finalité la rédaction de dictionnaires, car les dictionnaires ne doivent pas être rédigés. La lexicographie a pour finalité la construction de modèles génériques des lexiques des langues, modèles dont peuvent être dérivés de façon automatique et systématique, entre autres, des dictionnaires. La lexicographie doit être une lexicographie des dictionnaires virtuels.

Nous ne sommes pas les premiers à avancer une telle idée. On peut ainsi mentionner Sue Atkins qui, très explicitement, affirmait il y plus de quinze ans que les dictionnaires virtuels étaient le futur de la lexicographie (Atkins 1996). Malheureusement, à notre connaissance, seules des démonstrations du concept ont, jusqu'à ce jour, été réalisées (Spohr 2012). C'est la construction de véritables lexiques génériques, sources de dérivation de dictionnaires virtuels, qui permettra d'effectuer la transition vers la nouvelle lexicographie.

Parler de nouvelle lexicographie ne signifie pas que la tâche lexicographique doive profondément changer de nature. La « bonne » lexicographie existe, évolue et continuera d'évoluer de façon régulière, selon des principes maintenant relativement stables. Ce qui doit radicalement changer la lexicographie, c'est le fait de permettre au lexicographe d'exécuter une modélisation de chaque lexie-vedette à travers un tissage de liens entre lexies: là où le lexicographe d'aujourd'hui écrit un « mot » dans l'article d'un autre « mot », le lexicographe des dictionnaires virtuels doit établir un lien partant d'une lexie vers une autre lexie. Le lexicographe écrira peu de texte, parce que son activité le mènera avant tout à établir des connexions entre entités de natures diverses et, par là, à construire de façon explicite, et non plus implicite, le système lexical de la langue.

Remerciements

Merci à François Lareau, Veronika Lux-Pogodalla et Dorota Sikora pour leurs commentaires sur une version préliminaire de ce texte. Le projet RELIEF-RLF est financé par l'Agence de Mobilisation Économique de Lorraine (AMEL) et le Fonds Européen de Développement Régional (FEDER). Nous remercions chaleureusement toute l'équipe lexicographique du projet. Finalement, nous ne trouverons jamais tous les mots qu'il convient pour exprimer notre reconnaissance à Nabil Gader, développeur de l'éditeur lexicographique Dicet; sans lui, les travaux présentés ici seraient à jamais restés dans les cartons de notre imagination.

Références bibliographiques

Alonso et al. 2011 — *Alonso A., Millon C., Williams G.* Collocational networks and their application to an E-Advanced Learner's Dictionary of Verbs in Science (DicSci) // Proceedings of eLex 2011. Bled, (Slovénie), 2011. P. 12—22.

Atkins 1996 — *Atkins B. T. S.* Bilingual Dictionaries: Past, Present and Future // *Gellerstam M., Järborg J., Malmgren S.-G., Norén K., Rogström L., Papmehl C. R.* (eds.). Euralex'96 Proceedings. Gothenburg : Gothenburg University, Department of Swedish, 1996. P. 515—590.

Décary, Lapalme 1990 — *Décary M., Lapalme G.* An Editor for the Explanatory and Combinatory Dictionary of Contemporary French (DECFC) // Computational Linguistics. 1990. 16(3). P. 145—154.

Dendien, Pierrel 2003 — *Dendien J., Pierrel J.-M.* Le Trésor de la Langue Française informatisé: un exemple d'informatisation d'un dictionnaire de langue de référence // Traitement Automatique des Langues (T.a.l.). 2003. 44(2). P. 11—37.

Farias, Queiroz 2006 — *Farias P., Queiroz J.* Images, diagrams, and metaphors: Hypoicons in the context of Peirce's sixty-six-fold classification of signs // Semiotica. 2006. 162(1/4). P. 287—307.

Jousse, Polguère 2005 — *Jousse A.-L., Polguère A.* Le DiCo et sa version DiCouèbe. Document descriptif et manuel d'utilisation. Document technique. Département de linguistique et de traduction. Université de Montréal, 2005.

Lux-Pogodalla, Polguère 2011 — *Lux-Pogodalla V., Polguère A.* Construction of a French Lexical Network: Methodological Issues // Proceedings of the First International Workshop on Lexical Resources, WoLeR 2011. An ESSLLI 2011 Workshop. Ljubljana, 2011. P. 54—61.

Mel'čuk 1996 — *Mel'čuk I.* Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon // *Wanner L.* (ed.). Lexical Functions in Lexicography and Natural Language Processing. Language Companion Series 31. Amsterdam ; Philadelphia : John Benjamins, 1996. P. 37—102.

Mel'čuk 2006 — *Mel'čuk I.* Explanatory Combinatorial Dictionary // *Sica G.* (ed.). Open Problems in Linguistics and Lexicography. Monza : Polimetrica, 2006. P. 225—355.

Mel'čuk et al. 1995 — *Mel'čuk I., Clas A., Polguère A.* Introduction à la lexicologie explicative et combinatoire. Paris ; Louvain-la-Neuve : Duculot, 1995.
Mel'čuk et al. 1984—1999 — *Mel'čuk I. et al.* Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques. Vol. I—IV. Montréal : Les Presses de l'Université de Montréal, 1984—1999.
Mel'čuk, Polguère 2006 — *Mel'čuk I., Polguère A*. Dérivations sémantiques et collocations dans le DiCo/LAF // Langue française. 2006. 150. P. 66—83.
Mel'čuk, Polguère 2007 — *Mel'čuk I., Polguère A*. Lexique actif du français. L'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français // Champs linguistiques. Bruxelles: De Boeck & Larcier, 2007.
Mel'čuk, Zholkovsky 1984 — *Mel'čuk I., Zholkovsky A.* Explanatory Combinatorial Dictionary of Modern Russian. Semantico-syntactic Studies of Russian Vocabulary. Vienne : Wiener Slawistischer Almanach, 1984.
Polguère 1994 — *Polguère A.* The DiCo Editor User's Manual (ÉDiCo, version 1.0). Document technique. Department of English Language and Literature. National University of Singapore, 1994.
Polguère 2009 — *Polguère A.* Lexical systems : graph models of natural language lexicons // Language Resources and Evaluation. 2009. 43(1). P. 41—55.
Polguère 2011 — *Polguère A.* Classification sémantique des lexies fondée sur le paraphrasage // Cahiers de lexicologie. 2011. 98. P. 197—211.
Sagot, Fišer 2008 — *Sagot B., Fišer D.* Construction d'un wordnet libre du français à partir de ressources multilingues // Actes de TALN 2008. Avignon, 2008.
Sérasset 1998 — *Sérasset G.* Le projet NADIA-DEC: vers un Dictionnaire Explicatif et Combinatoire informatisé? // *Clas A., Mejri S., Baccouche T.* (eds.). La Mémoire des mots: Actes des cinquièmes Journées scientifiques du Réseau « Lexicologie, Terminologie, Traduction » de l'AUF. Tunis (Tunisie), 25—27 septembre 1997. Tunis ; Montréal : Serviced ; AUPELF-UREF, 1997. P. 149—159.
Spohr 2012 — *Spohr D.* Towards a Multifunctional Lexical Resource. Design and Implementation of a Graph-based Lexicon Model. Berlin ; Boston : De Gruyter, 2012.
Steinlin et al. 2005 — *Steinlin J., Kahane S., Polguère A.* Compiling a "classical" explanatory combinatorial lexicographic description into a relational database // Proceedings of the Second International Conference on the Meaning Text Theory (MTT'2005). M., 2005. P. 477—485.